JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

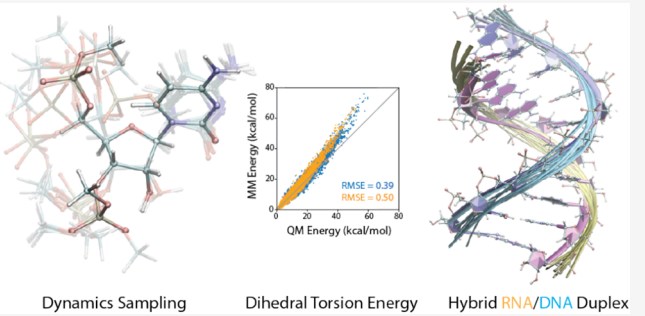
A Generalized Framework for Developing Dihedral Torsion Energy Terms in Nucleic Acids Force Fields

Chungwen Liang,* David Pekker, Alessio Valentini, and Swagatam Mukhopadhyay

Cite This: J. Chem. Theory Comput. 2025, 21, 7535−7549 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Accurate modeling of large-scale biomolecular systems depends on high-quality force field (FF) models. While atomistic nucleic acid FFs have improved over the past four decades, a universal framework for both natural and chemically modified nucleic acids across diverse environments remains lacking. In this work, we introduce a general methodology for developing torsional energy parameters that apply to all nucleic acid systems. Our approach simultaneously parametrizes key dihedral angles in nucleic acids critical to simulating their conformations at physiologically relevant temperatures and solvent environments. The resulting FF, Creyon25, achieves accuracy comparable to the latest AMBER and CHARMM models, but our framework in contrast is generalizable to chemical modifications (in linker, sugar and base). We validate Creyon25 across a wide range of RNA and DNA structures, including tetramers, tetraloops, and duplexes. We found that the Creyon25 RNA model accurately reproduces experimentally observed structures, although there is still room for improvement in the Creyon25 DNA model. This work represents a major step toward creating robust FFs for chemically modified nucleic acids, supporting the advancement of oligonucleotide therapies.



INTRODUCTION

Nucleic acids are essential for the regulation of numerous biological functions in all known forms of life. Their primary functions include serving as inheritable genetic material and facilitating the translation of genetic information into proteins. Beyond these fundamental functions, nucleic acid−based therapeutics have emerged as a powerful and rapidly evolving modality of medicine with a broad potential to address a wide range of human diseases.[1−5] Applications range from vaccine development,[6,7] to cancer treatment,[8−10] and precision medicines targeting rare diseases.[11−14] Consequently, understanding the structural, dynamic and electrostatic properties of natural and chemically modified nucleic acids is central to the advancement of the rapidly growing field of oligonucleotide therapies.[15]

Molecular dynamics (MD) simulation serves as a powerful tool, offering both high spatial and temporal resolution,[16] for probing those properties of nucleic acids. However, the accuracy of MD simulations is critically dependent on the quality of the energy models and the efficiency of conformational sampling. When modeling nucleic acids in explicit solvent environments, atomistic force fields (FFs) are typically used to study structural fluctuations (conformational changes) and dynamic processes occurring on nanosecond to microsecond time scales.[17,18] The first-generation atomistic RNA/DNA FFs were introduced in the early 1980s by the AMBER community.[19] Over the subsequent decades, numerous refinements to these FFs[20−31] have significantly improved the accuracy of nucleic acid simulations.[28,32] Notably, numerous studies have effectively demonstrated the capability of MD simulations using FF models to accurately predict duplex melting temperatures[33,34] and aptamer structures,[35,36] both of which are essential for advancing future development of oligonucleotide therapies. Nonetheless, the quality of current FF models remains far from ideal.[37] For instance, a recent study reported that while currently available nucleic acid FFs are good at predicting the conformation of RNA/RNA and DNA/DNA duplexes, none of them can accurately predict the conformation of *hybrid* RNA/DNA duplexes as observed in experiments.[38]

Past refinements of nucleic acid FFs have largely focused on improving the accuracy of torsion energies which play a critical role in determining nucleic acid conformations. However, these efforts often rely on assumptions that lack thorough justification. Specifically, (a) model compounds used for high-level quantum mechanical (QM) calculations are heavily

Received: April 15, 2025
Revised: July 23, 2025
Accepted: July 25, 2025
Published: July 31, 2025

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

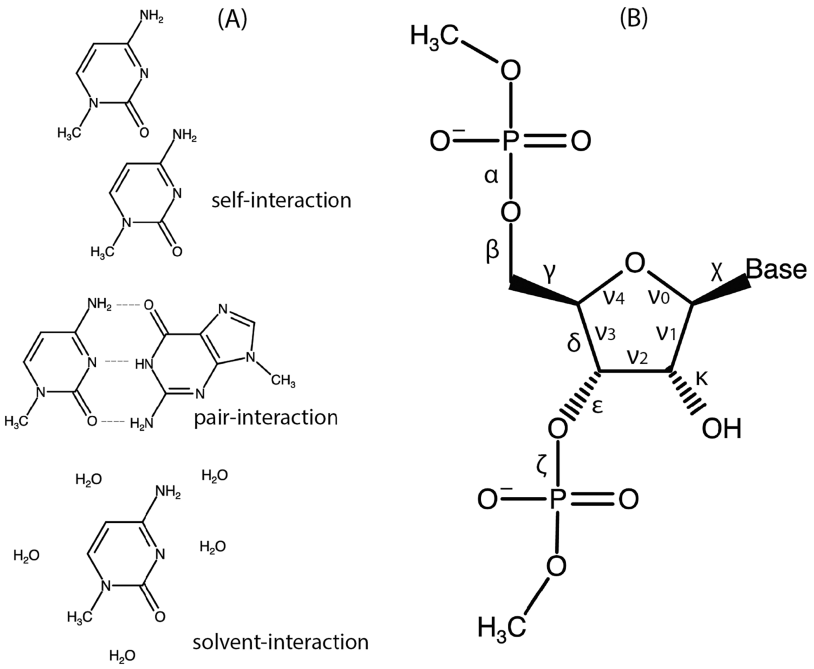


Figure 1. Model compounds for (A) base partial charges parametrization and (B) dihedral torsion energy development.

truncated,[28,31,39,40] resulting in the independent parametrization of torsions such as $\alpha/\beta/\gamma$ without accounting for coupling with ϵ/ζ or χ/κ (see Figure 1); (b) one-dimensional dihedral scans are generally performed without considering the collective behavior of torsion energies;[28,30,31,39−41] and (c) nonscanning dihedral torsion angles are assigned *ad hoc* values to avoid internal hydrogen bond formation.[28,40,42] These assumptions have led to decades of reparameterization and fine-tuning of individual or small sets of torsion energy terms in RNA/DNA FFs, often arbitrarily, to achieve desired accuracy when comparing to available experimental data.[28,31,39,41,42] Unmodified torsion parameters are then historically carried over from older FFs into newer ones, which could potentially compromise the quality of the updated parameter sets and result in inconsistencies. Furthermore, for chemically modified nucleic acids with diverse chemical modifications (in linker, sugar or base), this approach becomes impractical, as the scarcity of experimental data makes it unfeasible to individually refine each dihedral torsional term. A more systematic and generalized methodology is therefore required to address these challenges. Here, we offer such a methodology in this work. We note that the workflow used in the recently developed DNA force field Tumuc1[30] aims at addressing the issue of nonsystematically transferring older parameters to new models by reparameterizing all nucleobase partial charges and the key dihedral torsion energy terms. The latter are parametrized simultaneously using data from a collection of one-dimensional dihedral scans. This workflow could potentially be applicable to noncanonical nucleobases where experimental data is not available.

We design a more general methodology for developing torsion energy parameters that are critical for determining the structural and dynamic properties of nucleic acid systems. Importantly, our goal is not merely to enhance existing RNA or DNA FFs but to establish a framework that is broadly applicable to a wide range of chemistries beyond native RNA and DNA. As a preliminary step to apply this framework, we reparameterize the base partial charges to account for base−base and base-water interactions (Figure 1(a)). Our framework involves sampling of thousands of low-energy conformations of nucleotide model compounds (Figure 1(b)) and then fitting all important dihedral torsion energy terms concurrently. To validate the proposed framework, we have developed a new set of RNA and DNA FF models, named Creyon25, and tested it on several RNA, DNA, and hybrid systems. Creyon25 demonstrates comparable quality to the latest AMBER and CHARMM FFs, reasonably reproducing the structural properties of various RNA and DNA systems in agreement with experimental data. A major differentiation of Creyon25 lies in the reproduction of a hybrid RNA/DNA duplex structure, in which Creyon25 outperforms both latest AMBER and CHARMM FF models. These results showcase the potential of this framework in developing reliable FFs for chemically modified nucleic acids, which in turn holds a great promise for advancing oligonucleotide therapies.

The remainder of this paper is structured as follows. The Methods section describes the approach used to develop torsional energy parameters, detailing the underlying philosophy and rationale. This is followed by an overview of the computational procedures, including quantum chemistry calculations and MD simulations. In the Results section, we evaluate the performance of Creyon25 across various RNA and DNA systems, presenting detailed analyses to demonstrate its accuracy, effectiveness and expanded scope of application. The Discussion section highlights the strengths and limitations of our proposed framework, concluding with a summary of key findings and potential directions for future research.

METHODS

We describe our Methods as a general recipe for developing FFs for nucleic acids. To develop the torsion energy terms for nucleic acid systems, the following two-step workflow was adopted (illustrated in Figure 1): (a) The partial charges of the base atoms were determined by fitting the molecular mechanics (MM) energy (using FF models) to reproduce

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

the QM energy (as described below), while all the base-environment interactions were modeled, including base self-interactions (base stacking), base pair-interactions, and base-solvent interactions. (b) Four RNA model compounds (Figure 1(b)) were selected (PORC, PORU, PORG, and PORA): each model compound consists of a ribose sugar with a base (C, U, G, or A), flanked by methyl-capped 5′- and 3′-phosphate groups. The dihedral torsion energy terms were then obtained by fitting the MM energy to reproduce the QM energy. Note that the quality of the dihedral torsion energy parametrizations depends directly on the accuracy of the base partial charges, whose values are essential for accurately deriving the dihedral torsion energy in the specified nucleotide model compounds (PORC, PORU, PORG, and PORA). For DNA nucleotides, a similar procedure for obtaining the dihedral torsion energies was applied, using four DNA nucleotide model compounds (PODC, PODT, PODG, and PODA).

Base Partial Charge Fitting. For developing base partial charges of RNA, four *N*-methylated base molecules (C, U, G, and A) were chosen as model compounds. First, to evaluate the base self-interactions (stacking energy) between two identical base molecules (C−C, U−U, G−G, and A−A), *ab inito* MD simulations using the GFN2-xTB semiempirical method[43] in the gas phase were performed to sample their conformations. The simulations were performed at 600 K with the Nose−Hoover coupling method[44] (coupling time 10 fs). The total sampling time was 100 ps with 1.0 fs integration time. Structures were sampled at 100-step intervals, yielding 1000 configurations along the trajectory. Then, we performed a constrained optimization on each configuration at the Hartree−Fock (HF) level by fixing the N1 (in C and U) and N9 (in G and A) atoms. The purpose of the optimization was to relax the internal energy of the model compounds (e.g., bond lengths and angle bending), while restraining the relative geometries between two base model compounds. This allows us to evaluate the stacking energy between the base model compounds in different stacking geometries. Note that this sampling approach is fundamentally different from the traditional scanning method[28] for evaluating the base interaction energy. The QM single point energy calculation using RI-MP2[45]/aug-cc-pVTZ was then performed on each optimized configuration. All above QM calculations were performed using the ORCA 5.0.3 package.[46] Second, the same procedure was applied to evaluate the base pair-interactions (pairing energy) between two nonidentical bases of six systems (A−G, A−U, A−C, G−U, G−C, and U−C). Lastly, the base-solvent interaction was investigated as well. A base model compound solvated by five TIP4PD[47] water molecules was chosen as a model system. The hybrid quantum mechanics/molecular mechanics (QM/MM) MD samplings of four base-water systems (C−H_2O, U−H_2O, G−H_2O, and A−H_2O) were performed and followed the sampling procedure described above. Then, each configuration obtained by QM/MM MD samplings was optimized at the HF level, while all water oxygen atoms were constrained. The QM/MM single point energy calculation using RI-MP2/aug-cc-pVTZ was then performed on each optimized configuration. The total MM energies of the above 14 systems were then optimized and calculated by the Gromacs-2021 package[48] without periodic boundary conditions. The same restrained optimization procedure was applied (using the FF models) to ensure that the geometry of the MM conformers are similar to that of the QM ones. The initial base partial charges were taken from DESRES-FF.[28] The Nelder−Mead method in the Scipy package[49] was employed to optimize the base partial charges while minimizing the root-mean-square error (RMSE) between the QM (or QM/MM) and MM energy of above 14 systems simultaneously. Here, the RMSE is defined to be

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{n=0}^{N} (E_n^{\mathrm{QM}} - E_n^{\mathrm{MM}})^2 \cdot W_n}{\sum_{n=0}^{N} W_n}} \tag{1}$$

where E^{QM} is the single-point QM energy, E^{MM} is the MM energy, and N is the number of configurations. To improve the fit to low-energy conformations and reducing the contribution to the fit of high-energy conformations suggested by a previous work,[50] we introduced the weight factor

$$W_n = \mathrm{e}^{-(E_n^{\mathrm{QM}} - E_0^{\mathrm{QM}})/k_{\mathrm{B}}T} \tag{2}$$

where E_0^{QM} is the energy of the lowest energy conformer found in the ensemble, T is an effective temperature that should be treated as a hyperparameter of the fitting procedure, and k_{B} is the Boltzmann constant. For fitting the partial charges we set T = 2000 K—chosen to better reproduce the QM energy surface regions (0−10 kcal/mol) by the MM energy surfaces.

For the DNA base partial charges, the parameters of three bases (C, G, and A) are directly transferred from the RNA counterparts. The parameters of T are derived from U, by assigning standard methyl hydrogen charges in AMBER to H71/H72/H73 atoms, and selecting C5/C7 charges to ensure the net charge is zero. The evaluation of this assumption is discussed in the Supporting Information (SI) (Section 1.1). All the newly developed base partial charge values are summarized in the SI (Section 1.2).

Dihedral Torsion Energy Development. Our nucleotide model compounds for dihedral torsion energy development are shown in Figure 1(b). We chose a complete nucleotide monomer unit including methyl-capped 5′-end and 3′-end phosphate groups, a (deoxy)ribose, and a base (C, U, G, A, T). This choice allows us to evaluate the collective behavior of all important dihedral torsions in a complete nucleotide monomer unit.

To obtain a representative conformational ensemble of the model compounds, configurations were collected using two sampling schemes: meta-dynamics (MTD) sampling and CREST sampling. In the MTD sampling procedure, 5000 configurations of each model compound were obtained using the MTD[51] (with the GFN-FF generic force-field energy model[52]) sampling approach with the analytical linearized Poisson−Boltzmann (ALPB) implicit solvent model.[53] In the CREST sampling procedure,[54] the same energy (GFN-FF) and implicit solvent (ALPB) models were applied, and approximately 2000 additional configurations of each model compound were further collected. The input parameters of the MTD and CREST sampling procedures are provided in the SI (Section 1.3).

To derive a new set of dihedral torsion energy terms, $E_{\mathrm{dih}}^{\mathrm{MM}}(\{\phi\})$, the difference between the QM energy (E^{QM}) and the MM energy without the dihedral torsion contributions to be fitted ($E_{\mathrm{dih}=0}^{\mathrm{MM}}$) should be calculated

$$E_{\mathrm{dih}}^{\mathrm{MM}}(\{\phi\}) = E^{\mathrm{QM}} - E_{\mathrm{dih}=0}^{\mathrm{MM}} \tag{3}$$

The following procedure was applied to each configuration obtained by the two sampling procedures. First, each

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

configuration was optimized at the HF level in the implicit solvent (water environment) using the conductor-like polarizable continuum model (CPCM).[55] Then, a single point QM energy calculation at RI-MP2/aug-cc-pVTZ level in the gas phase was preformed to obtain E^{QM}. All QM calculations were performed using the ORCA 5.0.3 package.[46] To obtain the corresponding MM energy, each QM optimized configuration was reoptimized using a FF model based on the parameters (bonds, angles, nonparameterizing dihedrals, and nonbonded terms) taken from the literature[42] (but with the new base partial charges), while all the 13 (12 for DNA model compounds) dihedral torsion angles were constrained. Then, a single point MM energy calculation was performed in which the contribution of the 13 (12 for DNA model compounds) dihedral torsion energies were excluded to obtain $E^{MM}_{dih=0}$. All MM calculations were performed using the Gromacs-2021[48] package. Following the procedure outlined above, we obtained ∼7000 configurations (MTD and CREST) for each nucleotide model compound. The corresponding dihedral torsion angles $\{\phi\}$, E^{QM}, and $E^{MM}_{dih=0}$ were passed to the fitting procedure described below.

The new set of dihedral torsion energy was parametrized as a sum of multiple dihedral angles, where each dihedral torsion energy term i was defined by the following expansion

$$E^{MM}_{dih,j}(\phi_j) = \sum_{m=1}^{M} k_{j,m}[1 + \cos(m\phi_j - \theta_{j,m})] \quad (4)$$

where ϕ_j is a dihedral angle, $k_{j,m}$ and $\theta_{j,m}$ are the force constant and the phase angle for the mth term in the cosine expansion. In order to describe the dihedral torsion energy with sufficient accuracy, we chose $M = 4$ (each dihedral can be represented by a combination of four cosine functions). A total of 13 dihedral terms are fitted *simultaneously* for the RNA model compounds ($j \in \alpha, \beta, \gamma, \delta, \epsilon, \zeta, \chi, \kappa$, and ν_0 to ν_4). Twelve dihedral terms are fitted *simultaneously* for the DNA model compounds (same set of dihedrals with the exception of κ).

The dihedral torsion parameter fitting strategy involves recasting the dihedral torsion energy as a linear function of the fitting parameters, enabling the use of ridge regression for optimization. We started by formulating a new dihedral torsion energy, corresponding to the set of dihedral angles being fitted, in the following form

$$E^{MM}_{dih}(\{\phi\}) = \sum_{j} \sum_{m=1}^{M} a_{j,m} \sin(m\phi_j) + b_{j,m} \cos(m\phi_j) \quad (5)$$

where $a_{j,m}$ and $b_{j,m}$ are the fitting parameters. Equation 5 specifies the design matrix of a linear fitting problem in which each row corresponds to a different configuration, while eq 3 specifies the response vector, and eq 2 was used to specify the weight. We used the glmnet package[56] to perform ridge regression on the weighted linear fit problem. The glmnet package automatically optimizes the ridge parameter using cross-validation. After obtaining the fitting parameters, we discard the constant term (as it does not affect the dihedral forces) and map the fitting parameters $a_{i,m}$ and $b_{i,m}$ to parameters $k_{i,m}$ and $\theta_{i,m}$ which we use to generate the new dihedral torsion energy terms in a Gromacs compatible format.

In the dihedral torsion parameter fitting procedure, we again used eq 1 to evaluate the fitting quality. The effective temperatures of T = 1000, 1500, and 2000 K were employed for the PORC, PORA/PORG, and PORU systems, respectively, to achieve the most accurate reproduction of NMR ^{3}J-coupling constants obtained from RNA tetramer simulations (see Results section). For all DNA model compounds, an effective temperature of T = 2000 K was utilized uniformly. It should be emphasized that the selection of effective temperatures within the range of 1000 to 2000 K was intended to improve the agreement between MM energy surfaces and QM potential energy surfaces, particularly within the energy range of 0 to 10 kcal/mol.

Molecular Dynamics Simulations. Classical MD simulations were performed using the Gromacs-2021 package.[57] A wide variety of RNA and DNA systems described below were selected to benchmark the performance of Creyon25. For benchmarking purposes, we also selected the widely accepted AMBER RNA OL3[42] and DNA OL21[58,59] force field models to evaluate several RNA and DNA systems. The combination of RNA OL3 and DNA OL21 is referred to as AMBER14SB throughout this work.

RNA Tetramers. Four RNA tetramer systems were selected for validation: CCCC, UUUU, GACC, and AAAA. The initial configurations of these tetramers were prepared using the following procedure: First, a 10 ns MD simulation was performed in vacuum at 600 K, starting from an A-form conformation to allow the tetramer to adopt a fully extended state. The extended structure was then solvated in a 3.0 nm cubic box containing approximately 850 TIP4PD[60] water molecules and 0.15 M sodium chloride, with parameters taken from CHARMM22.[61] Next, the solvated system underwent equilibration at 300 K and 1 atm, using the Berendsen thermostat and barostat[62] for 10 ns. This was followed by a 10 ns canonical ensemble (NVT) simulation employing the stochastic velocity rescaling method[63] with an external heat bath set to 300 K (coupling time: 1 ps). After achieving equilibrium, we applied replica exchange with solute tempering (REST2)[64] to conduct replica exchange molecular dynamics (REMD) simulations. To enable REST2-based REMD, we compiled PLUMED-2.7.2[65] with Gromacs-2021.[57] Eight replicas were generated using equilibrated conformations from the previous NVT simulations, with temperatures ranging from 300 to 580 K and an acceptance ratio of 20−30% between neighboring replicas. Each replica was simulated for 1 μs, ensuring extensive sampling. For all subsequent analyses (unless otherwise stated), we used the trajectory from the 300 K replica, consisting of 40,000 snapshots after discarding the first 200 ns. To benchmark and compare the performance of Creyon25, we also modeled the tetramer systems using DESRES-FF[28] and CHARMM36.[66]

RNA Tetraloops. We selected three RNA tetraloop systems: gcUUCGgc (8mer), gggcGAGAgccu (12mer), and ggcacUUCGgugcc (14mer), where capitalized letters denote the loop regions. The initial configurations were obtained from experimental structures: X-ray for the 8mer (PDB: 1F7Y[67]) and NMR structures for the 12mer (PDB: 1ZIG[68]) and 14mer (PDB: 2KOC[69]). Notably, the gcUUCGgc tetraloop is a fragment extracted from an RNA-protein complex (PDB: 1F7Y). Each structure was solvated in a 5.0 nm cubic box containing approximately 4000 TIP4PD water molecules and 0.15 M sodium chloride. The equilibration procedure and simulation settings were identical to those used for the RNA tetramers. To evaluate tetraloop stability with Creyon25, we performed NVT (constant particle, temperature, and volume) simulations at 300 K for 2 μs per system. Additionally, two

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

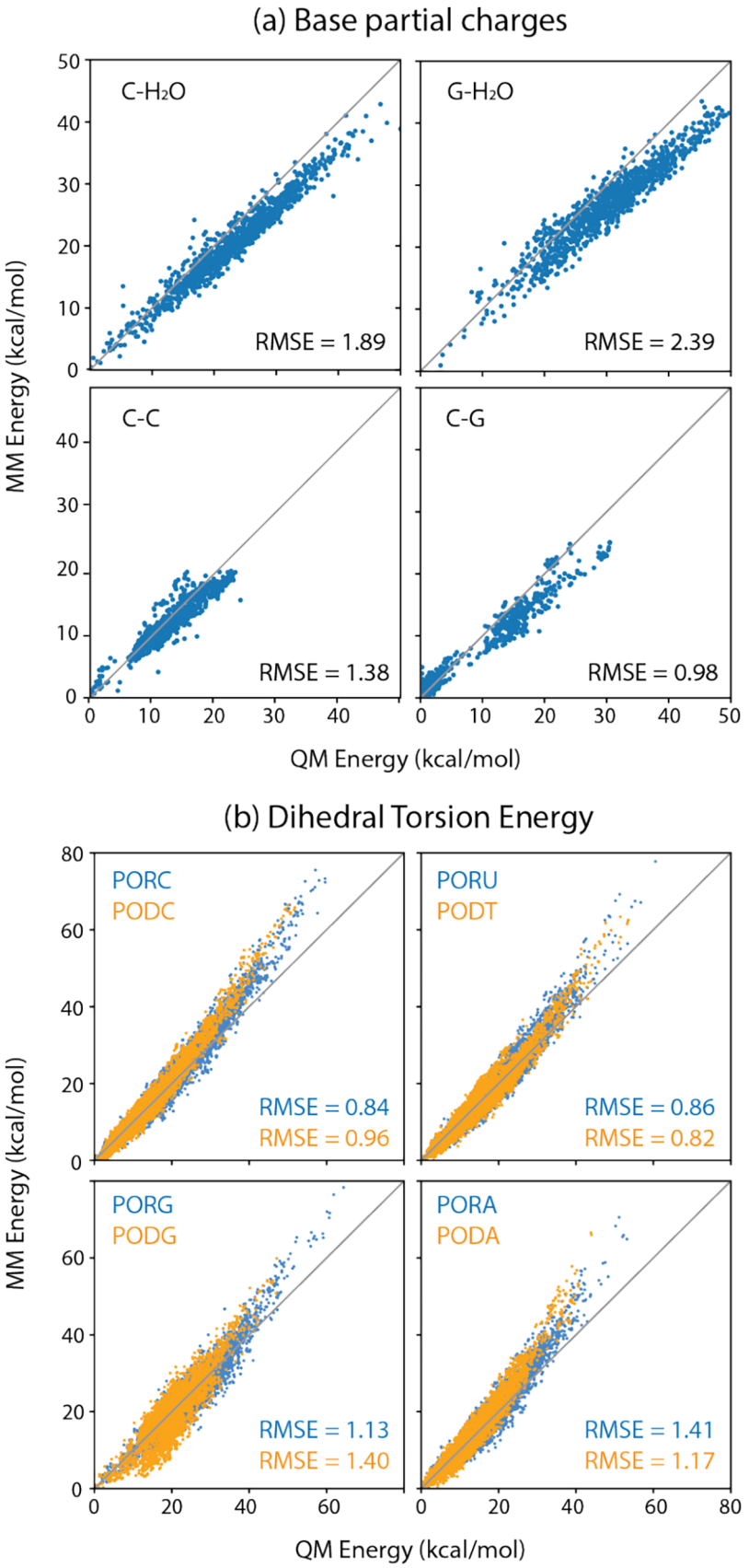


Figure 2. Correlation plots between the QM energy and the predicted MM energy for (a) base partial charge parametrization, and (b) dihedral torsion energy parametrization. Fitting RMSE (kcal/mol) is shown in the lower right corner of each plot.

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

extra simulations were conducted for the 8mer at 280 and 290 K to further assess its stability at lower temperatures.

RNA Duplexes. Two RNA duplex structures were selected for validation: GGUGAGG:CCUCACC(7mer) and GCAGAGAGCG:CGCUCUCUGC(10mer). The 7mer structure was derived from X-ray experiment (PDB: 3GVN[70]), while the 10mer was obtained from NMR spectroscopy (PDB: 2JXQ[71]). Both duplexes were solvated in 5.0 nm cubic boxes, each containing approximately 4000 TIP4PD water molecules and 0.15 M sodium chloride. The equilibration process and simulation parameters were identical to those used for RNA tetraloops. To assess duplex stability with Creyon25, we conducted NVT simulations at 300 K for 2 μs per system. For the benchmarking purpose, AMBER14SB (RNA OL3) force field was also selected to perform the 10mer simulation.

DNA Duplexes: A-DNA, B-DNA, and Z-DNA. We selected three common DNA duplex conformations: A-DNA, B-DNA, and Z-DNA. The initial structures were obtained from X-ray crystallography, specifically A-DNA (PDB: 1ZF9[72]), B-DNA (PDB: 1BNA[73]), and Z-DNA (PDB: 1ICK[74]). Each DNA duplex was solvated in an environment tailored to its structural characteristics: A-DNA was solvated in a 5.0 nm cubic box with a 40:60 water/ethanol mixture and 18 sodium ions to maintain charge neutrality. Ethanol FF parameters were taken from generalized AMBER force field (GAFF).[75] B-DNA was placed in a 6.0 nm cubic box containing approximately 6800 TIP4PD water molecules and 0.15 M sodium chloride. Z-DNA was solvated in a 5.0 nm cubic box with approximately 3700 TIP4PD water molecules and 2.0 M sodium chloride to replicate a high-salt environment. The equilibration process and simulation parameters followed the same protocol as those used for RNA tetraloops. To assess DNA duplex stability with Creyon25, we conducted NVT simulations at 300 K for 2 μs per system. For the benchmarking purpose, AMBER14SB (DNA OL21) force field was also selected to perform simulations of above three DNA systems.

Hybrid RNA/DNA Duplex. An RNA/DNA hybrid duplex was chosen for the validation as well. The initial configuration of the hybrid duplex was taken from the X-ray structure (PDB: 1PJG[76]). This structure was then solvated in 5.0 nm cubic boxes with ∼4000 TIP4PD water molecules and with 0.15 M sodium chloride. The equilibrium procedure and simulation setting were the same as that of RNA duplexes. To investigate the stability of the hybrid duplex with Creyon25, NVT simulation was performed at 300 K for 2 μs. For the benchmarking purpose, AMBER14SB force field was also selected to perform the above simulation.

Protein/RNA Complex. A recently resolved cryo-EM structure of the RNA duplex bound to the Ago2 protein (PDB: 9CMP[77]) provides a valuable model for studying siRNA-Ago2 interactions. We selected this system to benchmark the compatibility of Creyon25 with AMBER protein FFs. The Ago2 protein was modeled using the AMBER99SB-ILDN FF.[78] The initial configuration of the Ago2/RNA complex was obtained from the cryo-EM structure, with missing protein residues reconstructed using Swiss-Model homology modeling.[79] The magnesium ion at the reaction site was removed from the original structure. The system was then solvated in a 12.0 nm cubic box, containing approximately 52,000 TIP4PD water molecules and 0.15 M sodium chloride. The equilibration process consisted of the following steps: A 100 ns NPT simulation at 300 K was conducted with strong position restraints (1000 kJ/mol·nm^2) on all heavy atoms of the resolved protein residues and RNA, excluding residues reconstructed via homology modeling. Another 100 ns NPT simulation was performed, this time applying positional restraints only to the C_α atoms of the protein. A 100 ns NVT equilibrium simulation was then carried out at 300 K without any restraints. At this stage, the protein-RNA complex was considered equilibrated. After equilibration, a 1 μs production run was conducted at 300 K to evaluate the system's behavior with Creyon25 and the AMBER protein FF.

DNA G-Quadruplex. We selected a DNA G-quadruplex for the final validation. The initial structure was obtained from X-ray crystallography (PDB: 1JRN[80]). The G-quadruplex was solvated in a 5.0 nm cubic box containing approximately 4000 TIP4PD water molecules and 0.15 M potassium chloride. The equilibration process and simulation settings were identical to those used for RNA tetraloops. To evaluate the stability of the DNA G-quadruplex with Creyon25, we performed two constant-temperature simulations at 280 and 300 K for 2 μs each. Light positional restraints (100 kJ/mol·nm^2) were applied to the five central coordinated potassium ions to maintain structural integrity throughout the simulations. It is observed that in simulations using Creyon25, ions rapidly escape from their coordination sites (within 100 ns) in the absence of restraints, leading to significant distortion of the G-quadruplex structure. For the benchmarking purpose, AMBER14SB (DNA OL21) force field was also selected to perform simulations of G-quadruplex.

RESULTS

Force Field Parameterization. As mentioned in the Method section, base partial charges of model compounds are first parametrized simultaneously for 14 systems. Figure 2(a) shows the correlation plots between and QM energy and the predicted MM energy of four systems (C−H_2O, G−H_2O, C−C, and C−G). The effective temperature weighted RMSE (T = 2000 K) of each system was calculated using eq 1.

Correlation plots of other systems are shown in the SI (Section 1.1). The 13 dihedral torsion energy terms were parametrized simultaneously for RNA and model compounds (12 dihedral torsion energy terms for DNA model compounds). Fitting results for RNA (PORC, PORU, PORG, and PORA) and DNA (PODC, PODT, PODG, and PODA) model compounds are shown in Figure 2(b). The effective temperature weighted RMSE of each system is calculated using eq 1. The details of dihedral torsion energy surfaces and parameters of all model compounds are shown in the SI (Sections 1.4 and 1.5).

MD Simulations of Nucleic Acid Systems for Validation. *RNA Tetramers.* NMR ^{3}J-coupling experiments of short oligonucleotides provide structural information on the dihedral torsion angle distribution, which are often used to validate FF models.[81] We first performed REMD simulations of four RNA tetramers (CCCC, UUUU, GACC, and AAAA) to extensively sampling their configurations and calculated the NMR ^{3}J-coupling constants related to dihedrals using the Karplus equation.[81] We compared the qualities of three FF models: DESRES-FF, CHARMM36, and Creyon25. The results are summarized in Figure 3. More details of J-coupling constants for individual dihedral are presented in the SI (see Tables S5 and S6 in Section 2.1). By comparing the overall RMSEs, it shows that the quality of Creyon25 is in between DESRES-FF and CHARMM36. This indicates that Creyon25 reasonably reproduces the RNA tetramer conformational

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

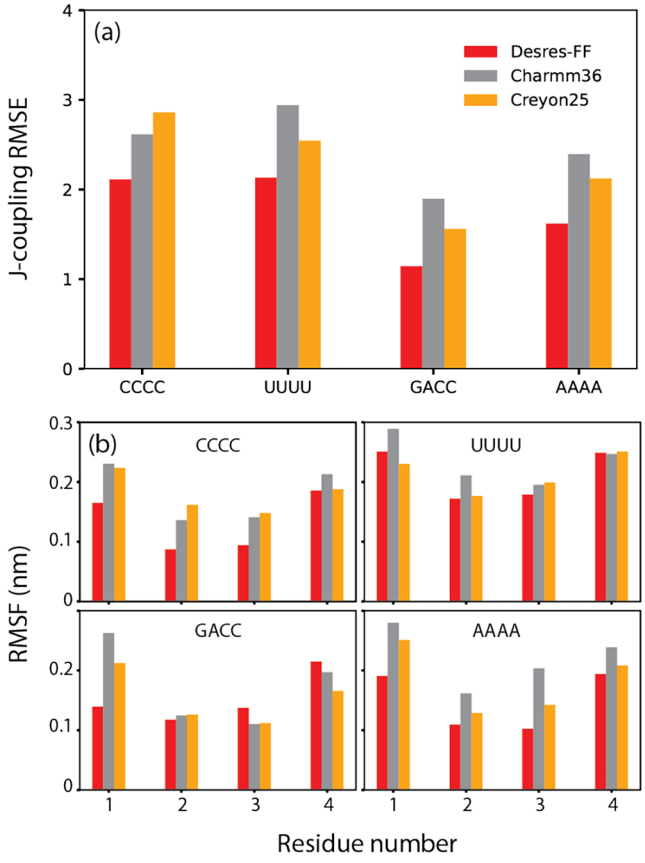


Figure 3. (a) NMR ^{3}J-coupling RMSE of four tetramers. (b) The RMSF of tetramers.

ensembles in solution (Figure 3(a)). The flexibility of tetramers was also compared by calculating the root-mean-square-fluctuation (RMSF) shown in (Figure 3(b)). It can be seen that, in general, DESRES-FF predicts slightly more rigid structures, and CHARMM36 predicts slightly more flexible ones, while Creyon25 is in between. A particularly interesting observation emerged from the UUUU tetramer simulations: both DESRES-FF and CHARMM36 predominantly predicted C3′-endo sugar puckering, whereas experimental data suggests a more balanced C2′- and C3′-endo population, based on sugar puckering-related dihedral angles (ν_1, ν_2, and ν_3). This discrepancy has also been reported in previous studies.[28] Notably, Creyon25 captured a more balanced distribution of C2′- and C3′-endo conformations, demonstrating its ability to accurately model sequence-dependent variations in sugar puckering.

RNA Tetraloops. Reproduction of RNA tetraloop structures using MD simulation is a crucial step for FF validation. We performed MD simulations of three RNA tetraloops with different sequences and lengths to further investigate the quality of Creyon25. To trace how the simulated structures deviate from the crystal and NMR ones, the backbone (O5′−C5′−C4′−C3′−O3−P) root-mean-square deviation (RMSD) as a function of simulation time for each system was calculated and shown in Figure 4. Note that terminal nucleotides were neglected for the RMSD calculations, due to their high flexibility in solution. It can be seen that for the 12mer and 14mer tetraloops, their backbone RMSDs with respect to the initial structures are below 0.25 nm within 2 μs simulation time, which indicates that Creyon25 well stabilize those tetraloop folded states found in NMR experiments. In contrast, for the 8mer, the backbone RMSD in the 300 K simulation rises sharply beyond 0.4 nm, reaching 0.6 nm after 200 ns, indicating that a large conformational change happened. This suggests that the short tetraloop, which has been experimentally observed to adopt a folded state in solution,[82] is not well-stabilized by Creyon25 at 300 K. This indicates that the folded state might not be the most dominant configuration at 300 K (approaching to its melting temperature), and the unfolding/refolding processes cannot be observed within short constant room temperature simulations. However, in the 280 and 290 K simulations, the RMSD remains below 0.25 nm, indicating that Creyon25 is capable of maintaining its folded state at temperatures below 290 K within 2 μs simulation.

RNA Duplexes. Two RNA duplexes were selected to further benchmark the Creyon25 RNA FF. Following the approach used for the tetraloop systems, we computed backbone RMSDs as a function of simulation time, as shown in Figure 5. To ensure a reliable assessment, terminal nucleotides were excluded from the RMSD calculations due to their high flexibility in solution. The results indicate that, for both duplexes, backbone RMSDs remain below 0.25 nm throughout the 2 μs simulation, demonstrating that Creyon25 accurately preserves the structural integrity of both X-ray and NMR-resolved duplex structures. Detailed duplex structural information on 7mer and 10mer are summarized in the SI (Section 2.2, Tables S6 and S7).

DNA Duplexes. To assess the accuracy of the Creyon25 DNA FF, we selected three commonly observed DNA conformations: A-DNA, B-DNA, and Z-DNA. Figure 6 presents the overlaid MD snapshots and crystal structures for these systems, along with their corresponding RMSD curves. In the B-DNA simulation with Creyon25, we observed a higher RMSD (~0.4 nm) compared to previous theoretical studies using recently developed DNA FFs, such as Tumuc1 (RMSD ~ 0.15 nm[30]) and DESRES-FF (RMSD ~ 0.25 nm[31]). Meanwhile, in simulations of A-DNA in an ethanol/water mixture and Z-DNA in a high-salt environment, RMSD values remained below 0.3 and 0.15 nm, respectively. We further compared the above simulations with those using AMBER14SB. Analysis of backbone RMSD and χ dihedral distributions (SI, Section 2.3) indicates that AMBER14SB performs better for B-DNA, while Creyon25 is more suitable for simulating A-DNA. Both force fields perform well for Z-DNA. These results suggest that Creyon25 accurately reproduces A-DNA and Z-DNA duplex conformations but is suboptimal for B-DNA. Detailed duplex structural information on A-DNA, B-DNA, and Z-DNA are summarized in the SI (Section 2.2, Tables S9−S11).

Accurately reproducing the BI/BII state population of B-DNA is a critical benchmark for assessing the quality of DNA FFs. It is well-known that the B-DNA backbone can adopt two distinct conformations: BI and BII, defined by the difference between the ϵ and ζ dihedral angles (BI: $\epsilon - \zeta < 0$ and BII: $\epsilon - \zeta > 0$).[83] In the B-DNA simulation with Creyon25, the calculated BI and BII populations were 83 and 17%, respectively. These values closely align with the BI/BII distribution (74/26%) reported in a recent theoretical study using the Tumuc1 model,[30] indicating that Creyon25 effectively captures the backbone conformational dynamics of B-DNA.

Hybrid RNA/DNA Duplex. To assess the compatibility of Creyon25 RNA and DNA models, we selected an RNA/DNA

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

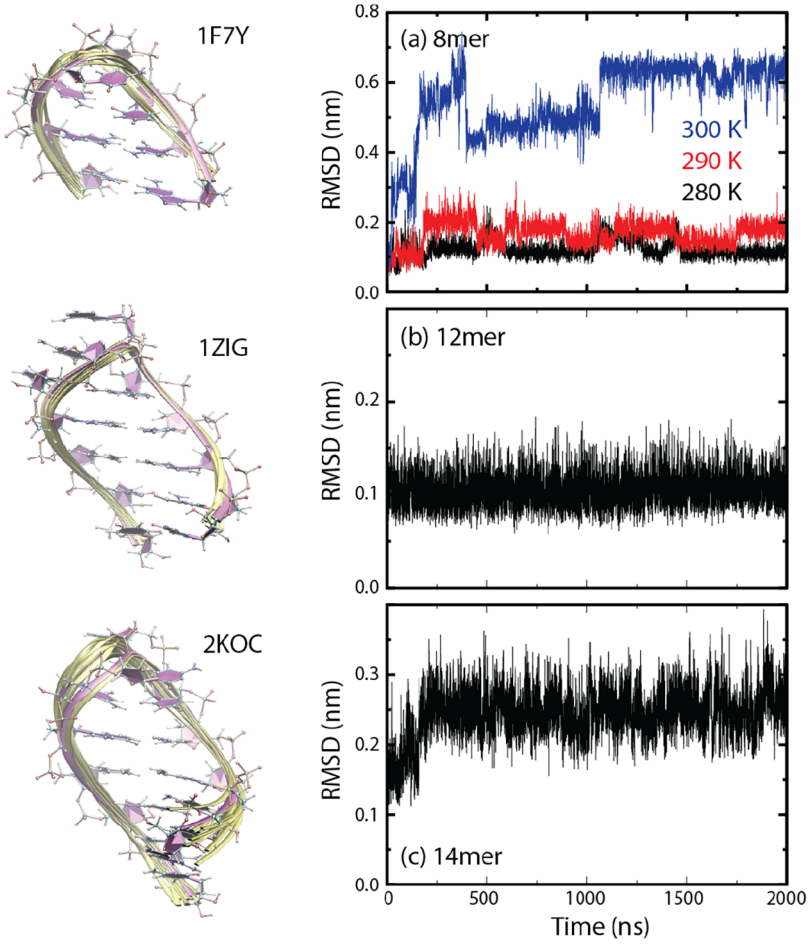


Figure 4. Backbone RMSD as a function of simulation time for RNA tetraloops: (a) 8mer, (b) 12mer, and (c) 14mer. To illustrate the conformational changes and stability of tetraloop systems, 10 MD snapshots (shown as yellow ribbons) along the full simulation trajectory for each system are overlapped with the X-ray or NMR structures (shown as pink ribbons).

hybrid duplex for validation. The simulation results are presented in Figure 7. The overlap of simulation snapshots with the crystal structure (Figure 7(a)) and the backbone RMSD over time (Figure 7(b)) demonstrate that Creyon25 effectively stabilizes the hybrid duplex, as resolved in X-ray experiments.[76] We further compared the above simulation with one using AMBER14SB. Analysis of backbone RMSD and χ dihedral distributions (SI, Section 2.3) indicates that Creyon25 is more suitable for simulating such hybrid RNA/DNA duplex. Detailed duplex structural information on the RNA/DNA hybrid duplex were summarized in the SI (Section 2.2, Table S12).

Beyond backbone conformation, DNA sugar puckering endo state is a key structural feature that differs depending on its hybridization state. Experimental evidence showed that in RNA/DNA hybrid duplexes, DNA sugar puckering predominantly adopts the C3′-endo state, whereas in homogeneous B-DNA duplexes, C2′-endo is the dominant conformation.[76,84] Accurately capturing this conformational shift is essential for validating the DNA puckering behavior in different environments. To evaluate this, we calculated the C2′- and C3′-endo state populations based on the backbone dihedral angle δ and the ring dihedral angle ν_2, as shown in Figure 7(c),(d). The C2′- and C3′-endo states are defined as following: C2′-endo: $\delta > 120°/\nu_2 < 0°$, and C3′-endo: $\delta < 120°/\nu_2 > 0°$. The results reveal that in the RNA/DNA hybrid duplex simulation, C3′-endo is strongly dominant (C2′/C3′ = 5%:94%), while in the B-DNA simulation, C2′-endo is much more prevalent (C2′/C3′ = 70%:26%). Notably, a similar C2′-endo population (∼76%) in a B-DNA duplex was also reported in a recent theoretical study.[30] These findings indicate that Creyon25 successfully reproduces the environment-dependent sugar puckering preferences of DNA in both RNA/DNA hybrid and B-DNA duplexes.

Protein/RNA Complex. The RNA-binding Ago2 protein is an ideal system for assessing the compatibility between nucleic acid and protein FFs. Figure 8 presents the simulation snapshots and backbone RMSD curve. Throughout the 1 μs simulation, the average backbone RMSD for RNA is ∼0.3 nm, demonstrating that Creyon25 maintains structural stability and is well compatible with AMBER protein FFs. We found that the larger RMSD of the protein backbone is primarily due to increased flexibility in the PAZ domain, consistent with observations from cryo-EM experiments[77] (Figure 8(c)). The presence of approximately 30 hydrogen bonds between the protein and RNA duplex indicates strong electrostatic interactions (Figure 8(d)), which are well captured by both the Creyon25 and AMBER force fields.

G-Quadruplex. To further assess the accuracy of the Creyon25 DNA model, we conducted MD simulations of a noncanonical DNA structure: G-quadruplex. Figure 9 presents the simulation snapshots and backbone RMSD curves. The

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

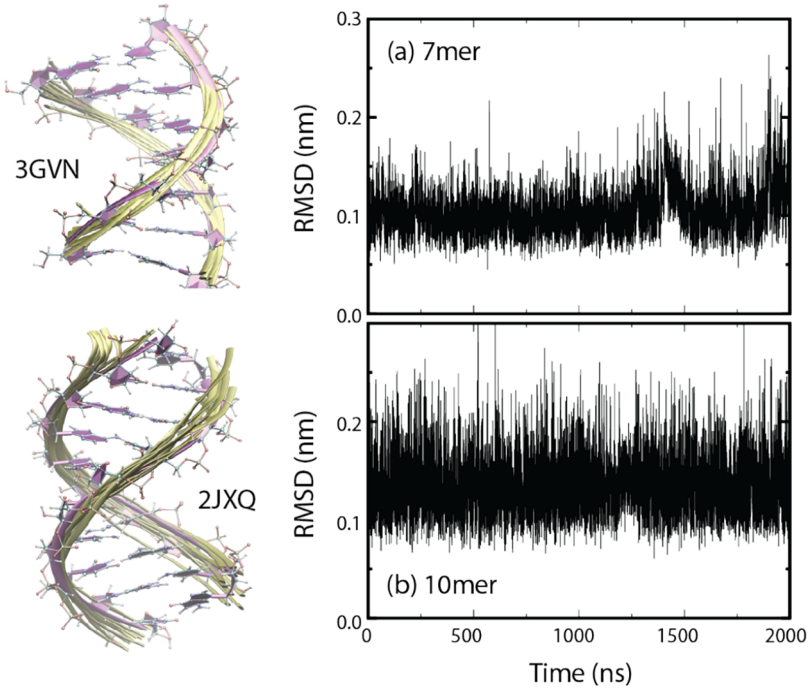


Figure 5. Backbone RMSD as a function of simulation time for RNA duplexes: (a) 7mer and (b) 10mer. To illustrate the conformational stability of the duplex systems, 10 MD snapshots (shown as yellow ribbons) along the full simulation trajectory for each system are overlapped with the X-ray or NMR structures (shown as pink ribbons).

results show that the MD structures deviate non-negligibly from the crystal structure even at low temperature (280 K), yielding an average RMSD of approximately 0.35 nm. At room temperature (300 K), the deviation becomes even more pronounced, with an RMSD of ∼0.6 nm. These findings suggest that the Creyon25 DNA model does not effectively capture the structural characteristics of G-quadruplexes with coordinated potassium ions. We further compared the above simulation with one performed using AMBER14SB. Analysis of backbone RMSD and χ dihedral distributions (SI, Section 2.3) indicates that AMBER14SB accurately captures the G-quadruplex structure, while there is room for improvement in the χ dihedral parameters of Creyon25 for accurately modeling G-quadruplex systems. A more detailed discussion of this discrepancy will be provided in the Discussion section.

DISCUSSION

The guiding philosophy of this work is to propose a generalized approach to develop FF models for a wide range of chemically modified nucleic acids without relying on arbitrary assumptions, *ad hoc* parameter fine-tuning, or other obtuse strategies. To the best of our knowledge, the parametrization of dihedral torsion energy terms in all existing nucleic acid FF models follows practices that are not justifiable in modern model fitting approaches. First, key dihedral torsion angles are typically parametrized independently. For example, $\alpha/\beta/\gamma$ are treated separately from ϵ/ζ, as well as χ and κ, often using heavily truncated model compounds for QM calculations. Next, one-dimensional scans are commonly performed for each key dihedral individually, with nonparameterized dihedral angles assigned arbitrary values to enforce desired geometries, such as avoiding internal hydrogen bond formation or enforcing sugar puckering in either C2′- or C3′-endo states. This could lead to the underestimation of several low populated but important states (e.g., the BII state in DNA). Furthermore, refinements tend to focus on improving only a subset of dihedral angles to meet quality requirements rather than reparameterizing all dihedral angles in a consistent and unified workflow. The framework proposed here seeks to overcome these limitations by offering a systematic and cohesive methodology for parametrizing dihedral torsion energy terms in nucleic acid FFs.

Several key improvements are introduced in our framework. First, complete nucleotide units, including both 5′- and 3′-end methyl-capped phosphate groups, a sugar, and a base, were chosen as model compounds. This comprehensive representation allows for the investigation of the collective behavior of all important dihedral torsions within a nucleotide unit. Instead of relying on traditional one-dimensional dihedral scans to collect model compound conformations for QM calculations, a dynamic sampling scheme is introduced. This approach captures a more realistic and physically meaningful conformational ensemble while avoiding artifacts caused by assigning *ad hoc* values to nonparameterized dihedral angles. Furthermore, all important torsion angles are parametrized simultaneously, eliminating issues related to treating parameters inconsistently, such as mixing old and new dihedral parameters. Another critical improvement is that the dihedral torsion parameters derived from this framework are specific to a given linker-sugar-base combination, rather than universally applicable. For instance, the $\alpha/\beta/\gamma/\delta/\epsilon/\zeta$ and κ/ν_n parameters are no longer assumed to be identical across nucleotide units with different bases, which is a common assumption in all current available nucleic acid FFs.

This specificity of dihedral torsion parameters enables a systematic application to develop FFs for modified nucleotides, such as those with phosphorothioate linkers, 2'-O-methyl sugars, or 7-deaza-guanosine bases—model compounds with new atom species or extra function groups would require highly specific parameters to correctly model their torsion

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

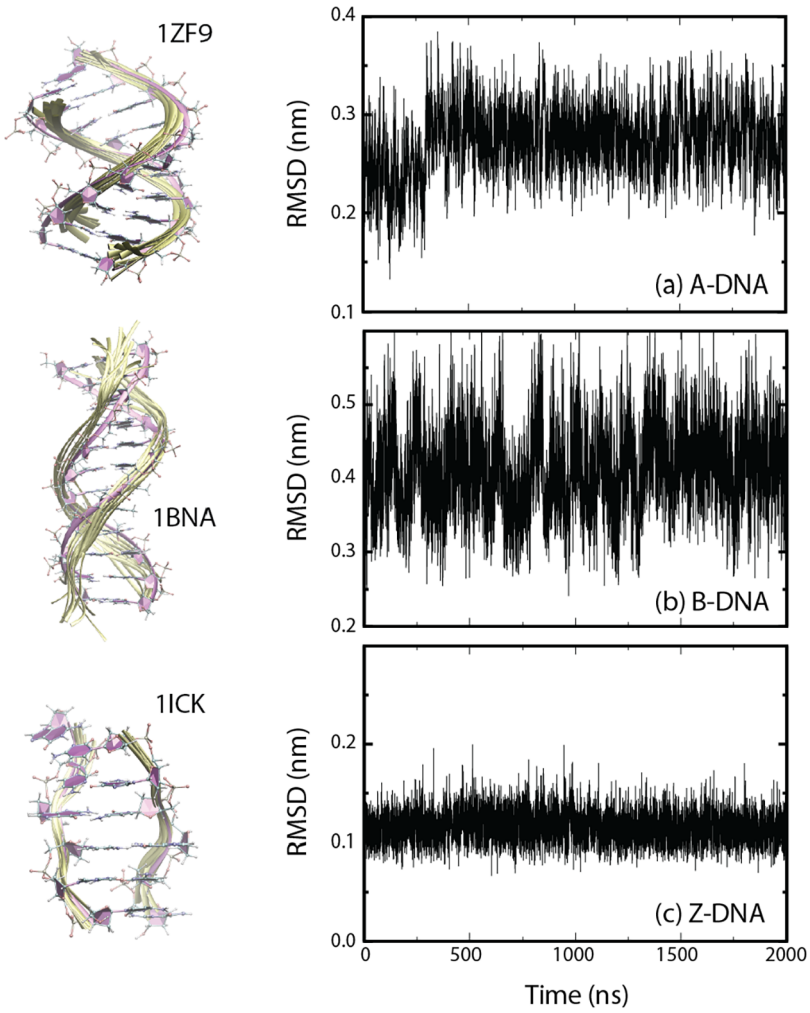


Figure 6. Backbone RMSD as a function of simulation time for (a) A-DNA, (b) B-DNA, and (c) Z-DNA duplexes. To illustrate the conformational stability of the duplex systems, 10 MD snapshots (shown as yellow ribbons) along the full simulation trajectory for each system are overlapped with the X-ray structures (shown as pink ribbons).

energies. However, this approach requires significantly higher computational cost, as parameters derived from this framework are not transferable across different nucleotide types. Consequently, each linker-sugar-base combination must be parametrized individually, making the process more resource-intensive. In addition to computational cost, we observed that the dihedral torsion parameters and the corresponding MD simulation outcomes are highly sensitive to the selection of energy models and conformational sampling schemes, as well as the hyperparameter T (effective temperature) in the fitting procedure. Several parametrization attempts employing alternative energy models (e.g., GFN2-xTB) and sampling strategies (e.g., MTD-only or CREST-only sampling) did not yield satisfactory results. A more detailed discussion of these unsuccessful parametrizations is provided in the SI (Section 3).

Among all the benchmark systems, we found that the RNA UUUU tetramer simulation with Creyon25 produced a sugar puckering endo state distribution that closely aligns with NMR experimental data, outperforming DESRES-FF and CHARMM36. The NMR findings (Table S5) indicate that the C2′- and C3′-endo states are similarly populated, as reflected in the closely matched ν_1 and ν_3 values. However, DESRES-FF and CHARMM36 exhibited a C3′-endo bias, making the UUUU tetramer appear more rigid than expected.[28] We hypothesize that this discrepancy arises from a simplified assumption in FF parametrization, where sugar puckering conformations were explicitly defined as either C2′- or C3′-endo during the dihedral torsion energy parametrization of nonpuckering-related dihedral angles (e.g., γ or χ). In the RNA tetraloop simulations with Creyon25, we observed that the 8mer unfolded significantly at 300 K within 200 ns but remained folded at temperatures below 290 K throughout the simulation. The simulation result is only qualitatively consistent with experimental studies on the 8mer UUCG tetraloop, where IR and UV spectroscopy data indicate a melting temperature of 54 °C (327 K).[82] Interestingly, a recent theoretical study reported that, even with finely tuned FF parameters, crystal UUCG tetraloop structures could not be stabilized in solution-phase MD simulations.[85] The study attributed this discrepancy to multiple interconnected FF inaccuracies that reinforce each other. To better reconcile simulation and experimental results, we propose employing enhanced sampling techniques, such as REMD or MTD, to investigate the folding mechanisms and melting temperatures of short 8mer RNA tetraloops. This approach could provide deeper insights into the structural dynamics of short tetraloops and help refine FF models accuracy.

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

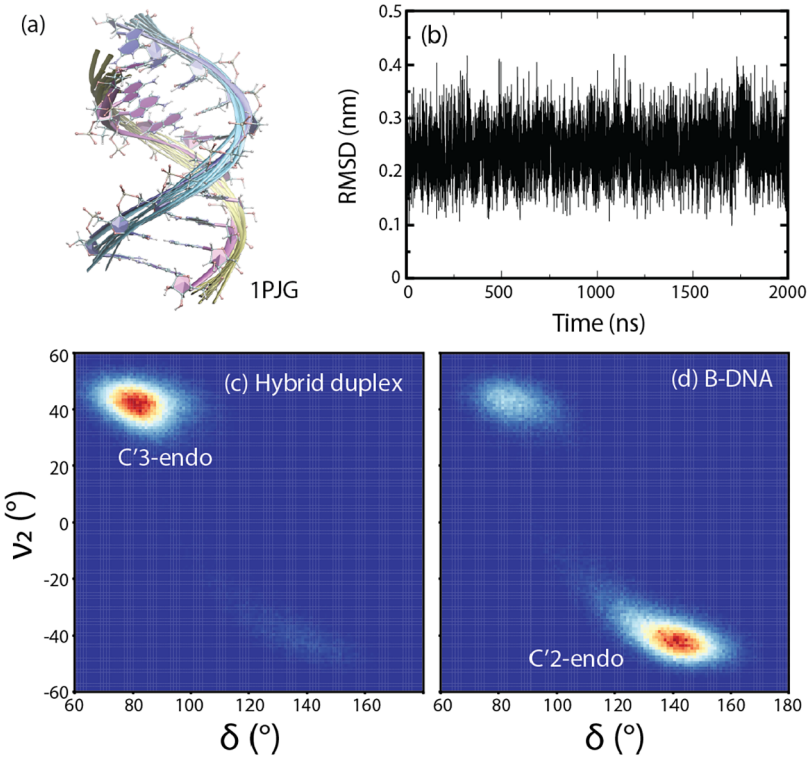


Figure 7. Simulation results of the RNA/DNA hybrid duplex: (a) an overlap of 10 simulation snapshots (RNA and DNA shown as yellow and blue ribbons, respectively) and the X-ray structure (RNA and DNA shown as pink and purple ribbons, respectively). (b) Backbone RMSD as a function of simulation time. The two-dimensional distributions of the δ and ν_2 dihedral angles of the DNA strand(s) in hybrid and B-DNA duplexes are shown in (c, d), respectively.

Current nucleic acid FF models face persistent challenges in accurately reproducing a diverse range of DNA conformations while also predicting their population distributions with accuracy. As discussed in the Introduction section, a recent theoretical study reported that none of the existing nucleic acid FF models can reliably model RNA/DNA hybrid duplexes, particularly in terms of sugar puckering endo state populations.[38] However, accurately simulating hybrid duplexes is essential for understanding how synthetic nucleic acids interact with RNA target strands, which is a crucial step toward *in silico* oligonucleotide design. With Creyon25, we demonstrated its ability to predict a reasonable BI/BII state population in B-DNA and accurately capture the transition between C2′- and C3′-endo states of the DNA sugar puckering, particularly in the shift from B-DNA to RNA/DNA hybrid duplexes. However, several areas require further refinement. One notable limitation is the higher RMSD values observed for B-DNA and the G-quadruplex found in the present work. Compared to AMBER14SB, the Creyon25 DNA model does not accurately reproduce the stable base configurations (as defined by the χ dihedral angle) in B-DNA and G-quadruplex structures, resulting in larger RMSD values relative to crystal structures. However, Creyon25 outperforms AMBER14SB in modeling A-DNA and RNA/DNA hybrid systems. We found that the χ dihedral torsion term in the Creyon25 DNA model biases DNA duplexes energetically toward the A-form rather than the B-form. A recent theoretical study[38] similarly highlighted the persistent inability of current force fields to describe A- and B-DNA simultaneously. Most DNA force fields show related deficiencies—such as imbalance in DNA/RNA hybrids and poor A-DNA representation—with no clear solution yet available. To better reproduce the diversity of DNA structures observed experimentally, further optimization of DNA dihedral parameters—particularly the χ angle—is necessary. To address these limitations, we recommend incorporating higher-level ab initio molecular dynamics (AIMD) sampling, such as Density Functional Theory (DFT) or Møller—Plesset second-order perturbation theory (MP2), in place of semiempirical methods for dihedral parametrization. Semiempirical models like GFN-FF and GFN2-xTB may introduce biases by overemphasizing low-populated conformations, leading to imbalanced parameter sets. In contrast, AIMD-based sampling can generate more reliable conformational ensembles of nucleotide model compounds, thereby enhancing the quality of force field parametrization and improving the accuracy of simulations across a broader range of nucleic acid structures.

A very recent theoretical study introduced a new tool, modXNA,[86] designed for parametrizing FF models of chemically modified nucleic acids that are compatible with the AMBER FF family. Notably, in the modXNA framework, partial charges for modified nucleic acid model compounds are derived using the AMBER restrained electrostatic potential (RESP) fitting procedure. However, the dihedral torsion parameters—which are critical for accurately modeling overall structural conformations—are directly adopted from existing RNA/DNA or GAFF parameter sets, without reparameterization. We hypothesize that by integrating modXNA with the Creyon25 framework introduced in this work, one can efficiently generate reliable FF models for chemically modified nucleic acids. This represents a crucial advancement toward *in silico* oligonucleotide design and development.

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

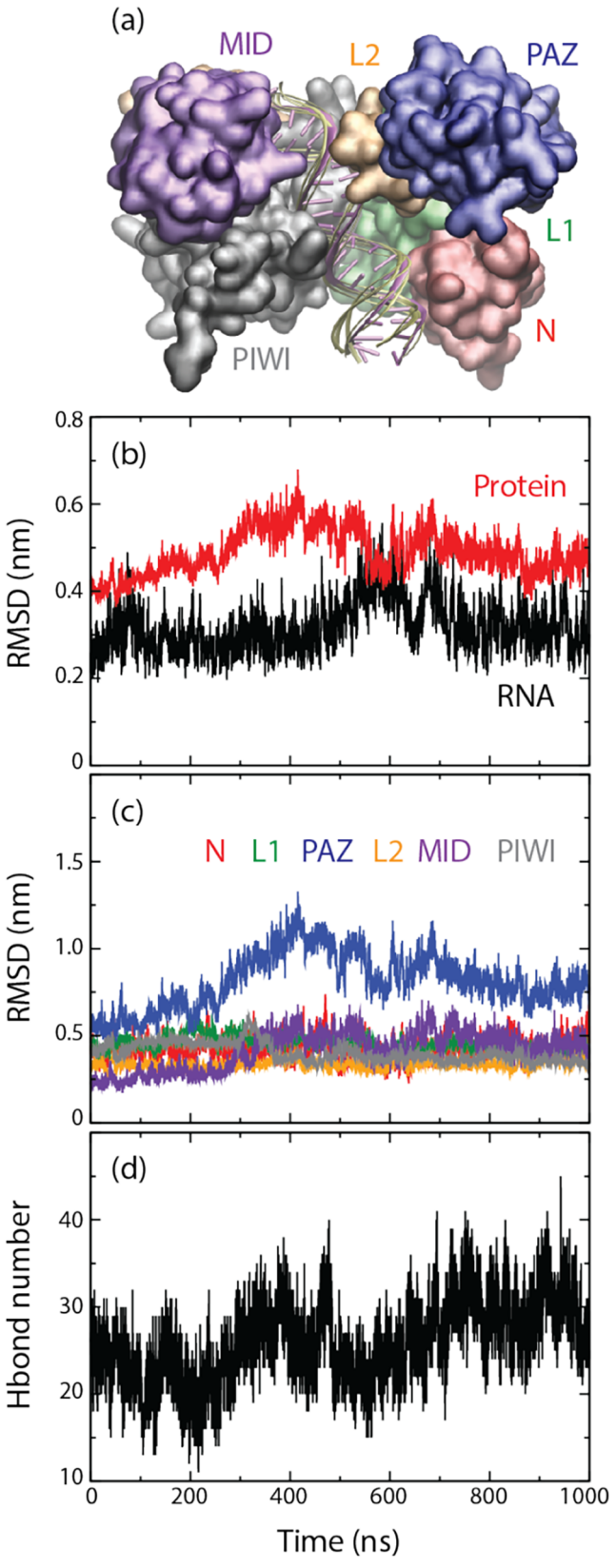


Figure 8. Simulation results of the Ago2/RNA complex: (a) an overlap of 10 RNA duplex snapshots (yellow ribbons) and the X-ray structure (pink ribbons). Protein X-ray structure is shown as color surface (b) the RNA and protein backbone RMSDs as a function of simulation time. (c) Protein backbone RMSDs of different domains. (d) Hydrogen bond number between RNA and protein.

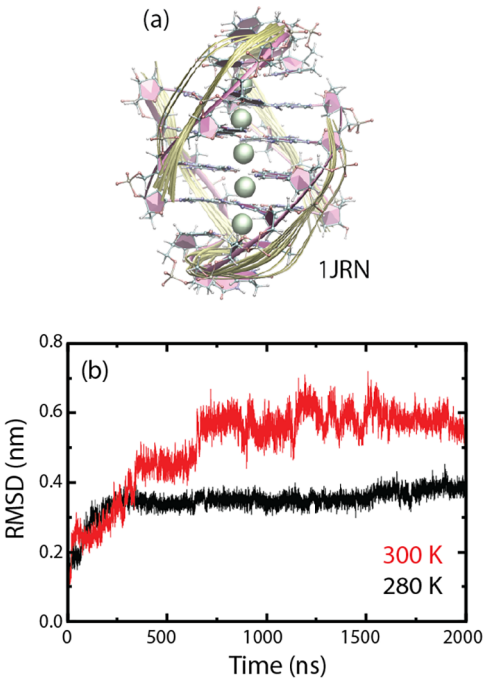


Figure 9. Simulation results of the G-quadruplex: (a) an overlap of 10 simulation snapshots (yellow ribbons) of the simulation at 280 K and the X-ray structure (pink ribbons). (b) The backbone RMSDs at 280 and 300 K as a function of simulation time.

CONCLUSIONS

In this work, we introduce a generalized framework for developing torsion energy parameters in nucleic acid FFs, which are essential for accurately predicting both secondary and tertiary structures of nucleic acids. Using this approach, we have developed a new set of RNA and DNA FF models, named Creyon25, which demonstrates performance comparable to the latest AMBER and CHARMM FFs. A key strength of this framework is its ability to simultaneously parametrize all critical dihedral angles while accounting for their collective behavior, eliminating the need for arbitrary assumptions about nonparameterized dihedral angles. We believe that adopting this generalized framework represents a major advancement in the development of FF models for modified nucleic acids. This progress will enhance the capabilities of molecular modeling and simulation, enabling more accurate studies of the structural, dynamic, and electrostatic properties of oligonucleotides in the near future.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.5c00602.

Correlation plots between the referened QM energy and the predicted MM energy of base model compounds for base partial charge development (Figure S1); dihedral torsion energy surfaces of PORC and PODC model compounds (Figure S2); dihedral torsion energy surfaces of PORU and PODT model compounds (Figure S3); dihedral torsion energy surfaces of PORG and PODG model compounds (Figure S4); dihedral torsion energy surfaces of PORA and PODA model compounds (Figure S5); comparison between Creyon25 and AMBER14SB on several duplex systems (Figure

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

S6); simulation results of the B-DNA duplex with the dihedral parameters derived from the GFN2-xTB energy model for conformational samplings (Figure S7); list of partial charges of adenine and cytosine bases (Table S1); list of partial charges of guanine and uracil bases (Table S2); list of partial charges of thymine base (Table S3); dihedral torsion energy parameters of RNA/DNA model compounds (Table S4); NMR ^{3}J-coupling constants of CCCC and UUUU tetramers (Table S5); NMR ^{3}J-coupling constants of GACC and AAAA tetramers (Table S6); RNA 7mer duplex structural information (Table S7); RNA 10mer duplex structural information (Table S8); A-DNA duplex structural information (Table S9); B-DNA duplex structural information (Table S10); Z-DNA duplex structural information (Table S11), and hybrid RNA/DNA duplex structural information (Table S12) (PDF)

AUTHOR INFORMATION

Corresponding Author

Chungwen Liang − *Creyon Bio Inc., Carlsbad, California 92010, United States;* orcid.org/0000-0001-9721-6411; Email: chungwen.liang@gmail.com

Authors

David Pekker − *Creyon Bio Inc., Carlsbad, California 92010, United States; Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States*

Alessio Valentini − *Creyon Bio Inc., Carlsbad, California 92010, United States*

Swagatam Mukhopadhyay − *Creyon Bio Inc., Carlsbad, California 92010, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.5c00602

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Giovanni Bussi for insightful discussion.

REFERENCES

(1) Cooper, T. A.; Wan, L.; Dreyfuss, G. RNA and disease. *Cell* **2009**, *136*, 777−793.

(2) Harries, L. W. RNA Biology Provides New Therapeutic Targets for Human Disease. *Front. Genet.* **2019**, *10*, No. 205.

(3) Scotti, M. M.; Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **2016**, *17*, 19−32.

(4) Wolin, S. L.; Lynne, E. M. Cellular RNA surveillance in health and disease. *Science* **2019**, *366*, 822−827.

(5) Adachi, H.; Hengesbach, M.; Yu, Y.-T.; Morais, P. From Antisense RNA to RNA Modification: Therapeutic Potential of RNA-Based Technologies. *Biomedicines* **2021**, *9*, No. 550.

(6) Batista-Duharte, A.; Sendra, L.; Herrero, M. J.; Téllez-Martínez, D.; Carlos, I. Z.; Aliño, S. F. Progress in the Use of Antisense Oligonucleotides for Vaccine Improvement. *Biomolecules* **2020**, *10*, No. 316.

(7) Oberemok, V. V.; Laikova, K. V.; Yurchenko, K. A.; Novikov, I. A.; Makalish, T. P.; Kubyshkin, A. V.; Andreeva, O. A.; Bilyk, A. I. Adjuvant Oligonucleotide Vaccine Increases Survival and Improves Lung Tissue Condition of B6.Cg-Tg (K18-ACE2)2 Transgenic Mice. *Sci. Pharm.* **2023**, *91*, No. 35.

(8) Xiong, H.; Veedu, R. N.; Diermeier, S. D. Recent Advances in Oligonucleotide Therapeutics in Oncology. *Int. J. Mol. Sci.* **2021**, *22*, No. 3295.

(9) Bartolucci, D.; Pession, A.; Hrelia, P.; Tonelli, R. Precision Anti-Cancer Medicines by Oligonucleotide Therapeutics in Clinical Research Targeting Undruggable Proteins and Non-Coding RNAs. *Pharmaceutics* **2022**, *14*, No. 1453.

(10) Dean, N. M.; Bennett, C. F. Antisense oligonucleotide-based therapeutics for cancer. *Oncogene* **2003**, *22*, 9087−9096.

(11) Aoki, Y.; Wood, M. J. Emerging Oligonucleotide Therapeutics for Rare Neuromuscular Diseases. *J. Neuromuscular Dis.* **2021**, *8*, 869−884.

(12) Lauffer, M. C.; van Roon-Mom, W.; Aartsma-Rus, A. Possibilities and limitations of antisense oligonucleotide therapies for the treatment of monogenic disorders. *Commun. Med.* **2024**, *4*, No. 6.

(13) Igarashi, J.; Niwa, Y.; Sugiyama, D. Research and Development of Oligonucleotide Therapeutics in Japan for Rare Diseases. *Future Rare Dis.* **2022**, *2*, No. FRD19.

(14) Kim, J.; Hu, C.; El Achkar, C. M.; et al. Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. *N. Engl. J. Med.* **2019**, *381*, 1644−1652.

(15) Moumné, L.; Marie, A.-C.; Crouvezier, N. Oligonucleotide Therapeutics: From Discovery and Development to Patentability. *Pharmaceutics* **2022**, *14*, No. 260.

(16) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646−652.

(17) Cheatham, T. E., III; Kollman, P. A. Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435−471.

(18) Vangaveti, S.; Ranganathan, S. V.; Chen, A. A. Advances in RNA molecular dynamics: a simulator's guide to RNA force fields. *WIREs RNA* **2017**, *8*, No. e1396.

(19) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(20) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(21) Aduri, R.; Psciuk, B. T.; Saro, P.; Taniga, H.; Schlegel, H. B.; SantaLucia, J. AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. *J. Chem. Theory Comput.* **2007**, *3*, 1464−1475.

(22) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of /Conformers. *Biophys. J.* **2007**, *92*, 3817−3829.

(23) Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T. E.; Jurečka, P.; Šponer, J. Reference Simulations of Noncanonical Nucleic Acids with Different ?Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8*, 2506−2520.

(24) Condon, D. E.; Yildirim, I.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Turner, D. H. Optimization of an AMBER Force Field for the Artificial Nucleic Acid, LNA, and Benchmarking with NMR of L(CAAU). *J. Phys. Chem. B* **2014**, *118*, 1216−1228.

(25) Gil-Ley, A.; Bottaro, S.; Bussi, G. Empirical Corrections to the Amber RNA Force Field with Target Metadynamics. *J. Chem. Theory Comput.* **2016**, *12*, 2790−2798.

(26) Darré, L.; Ivani, I.; Dans, P. D.; Gómez, H.; Hospital, A.; Orozco, M. Small Details Matter: The 2'-Hydroxyl as a Conformational Switch in RNA. *J. Am. Chem. Soc.* **2016**, *138*, 16355−16363.

(27) Aytenfisu, A. H.; Spasic, A.; Grossfield, A.; Stern, H. A.; Mathews, D. H. Revised RNA Dihedral Parameters for the Amber Force Field Improve RNA Molecular Dynamics. *J. Chem. Theory Comput.* **2017**, *13*, 900−915.

(28) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E1346−E1355.
(29) Kührová, P.; Mlýnský, V.; Zgarbová, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Šponer, J.; Banáš, P. Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J. Chem. Theory Comput.* **2019**, *15*, 3288−3305.
(30) Liebl, K.; Zacharias, M. Tumuc1: ANew Accurate DNA Force Field Consistent with High-Level Quantum Chemistry. *J. Chem. Theory Comput.* **2021**, *17*, 7096−7105.
(31) Tucker, M. R.; Piana, S.; Tan, D.; LeVine, M. V.; Shaw, D. E. Development of Force Field Parameters for the Simulation of Single- and Double-Stranded DNA Molecules and DNA-Protein Complexes. *J. Phys. Chem. B* **2022**, *126*, 4442−4457.
(32) Robertson, M. J.; Qian, Y.; Robinson, M. C.; Tirado-Rives, J.; Jorgensen, W. L. Development and Testing of the OPLS-AA/M Force Field for RNA. *J. Chem. Theory Comput.* **2019**, *15*, 2734−2742.
(33) Lomzov, A. A.; Vorobjev, Y. N.; Pyshnyi, D. V. Evaluation of the Gibbs Free Energy Changes and Melting Temperatures of DNA/DNA Duplexes Using Hybridization Enthalpy Calculated by Molecular Dynamics Simulation. *J. Phys. Chem. B* **2015**, *119*, 15221−15234.
(34) Wong, K.-Y.; Pettitt, B. M. The Pathway of Oligomeric DNA Melting Investigated by Molecular Dynamics Simulations. *Biophys. J.* **2008**, *95*, 5618−5626.
(35) Ropii, B.; Bethasari, M.; Anshori, I.; Koesoema, A. P.; Shalannanda, W.; Satriawan, A.; Setianingsih, C.; Akbar, M. R.; Aditama, R. The assessment of molecular dynamics results of three-dimensional RNA aptamer structure prediction. *PLoS One* **2023**, *18*, No. e0288684.
(36) Autiero, I.; Vitagliano, L. Enhanced molecular dynamic simulation studies unravel long-range effects caused by sequence variations and partner binding in RNA aptamers. *Mol. Ther.–Nucleic Acids* **2023**, *34*, No. 102039.
(37) Liebl, K.; Zacharias, M. The development of nucleic acids force fields: From an unchallenged past to a competitive future. *Biophys. J.* **2023**, *122*, 2841−2851.
(38) Knappeová, B.; Mlýnský, V.; Pykal, M.; Šponer, J.; Banáš, P.; Otyepka, M.; Krepl, M. Comprehensive Assessment of Force-Field Performance in Molecular Dynamics Simulations of DNA/RNA Hybrid Duplexes. *J. Chem. Theory Comput.* **2024**, *20*, 6917−6929.
(39) Yildirim, I.; Kennedy, S. D.; Stern, H. A.; Hart, J. M.; Kierzek, R.; Turner, D. H. Revision of AMBER Torsional Parameters for RNA Improves Free Energy Predictions for Tetramer Duplexes with GC and iGiC Base Pairs. *J. Chem. Theory Comput.* **2012**, *8*, 172−181.
(40) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory Comput.* **2018**, *14*, 2084−2108.
(41) Zgarbová, M.; Luque, F. J.; Šponer, J.; Cheatham, T. E. I.; Otyepka, M.; Jurečka, P. Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013**, *9*, 2339−2354.
(42) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E. I.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886−2902.
(43) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB−An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.
(44) Braga, C.; Travis, K. P. A configurational temperature Nosé-Hoover thermostat. *J. Chem. Phys.* **2005**, *123*, No. 134101.
(45) Ishikawa, T.; Kuwata, K. RI-MP2 Gradient Calculation of Large Molecules Using the Fragment Molecular Orbital Method. *J. Phys. Chem. Lett.* **2012**, *3*, 375−379.
(46) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **2020**, *152*, No. 224108.
(47) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.
(48) Lindahl, E.; Hess, B.; van der Spoel, D. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306−317.
(49) Virtanen, P.; Gommers, R.; Oliphant, T. E.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261−272.
(50) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 3499−3509.
(51) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847−2862.
(52) Spicher, S.; Grimme, S. Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems. *Angew. Chem., Int. Ed.* **2020**, *59*, 15665−15673.
(53) Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S. Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods. *J. Chem. Theory Comput* **2021**, *17*, 4250−4261.
(54) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169−7192.
(55) Takano, Y.; Houk, K. N. Benchmarking the Conductor-like Polarizable Continuum Model (CPCM) for Aqueous Solvation Free Energies of Neutral and Ionic Organic Molecules. *J. Chem. Theory Comput.* **2005**, *1*, 70−77.
(56) Friedman, J.; Tibshirani, R.; Hastie, T. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Software* **2010**, *33*, 1−22.
(57) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(58) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E. r.; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput* **2015**, *11*, 5723−5736.
(59) Zgarbová, M.; Šponer, J.; Jurečka, P. Z-DNA as a Touchstone for Additive Empirical Force Fields and a Refinement of the Alpha/Gamma DNA Torsions for AMBER. *J. Chem. Theory Comput* **2021**, *17*, 6292−6301.
(60) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.
(61) MacKerell, A. D. J.; Bashford, D.; Bellott, M.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(62) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(63) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
(64) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, No. 13749.
(65) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(66) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D., Jr. Impact of 2'-hydroxyl sampling on the conformational properties of

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549

RNA: Update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929−1943.
(67) Ennifar, E.; Nikulin, A.; Tishchenko, S.; Serganov, A.; Nevskaya, N.; Garber, M.; Ehresmann, B.; Ehresmann, C.; Nikonov, S.; Dumas, P. The crystal structure of UUCG tetraloop11Edited by J Doudna. *J. Mol. Biol.* **2000**, *304*, 35−42.
(68) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. A Network of Heterogeneous Hydrogen Bonds in GNRA Tetraloops. *J. Mol. Biol.* **1996**, *264*, 968−980.
(69) Nozinovic, S.; Fürtig, B.; Jonker, H. R. A.; Richter, C.; Schwalbe, H. High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res.* **2010**, *38*, 683−694.
(70) Eichert, A.; Fürste, J. P.; Schreiber, A.; Perbandt, M.; Betzel, C.; Erdmann, V. A.; Förster, C. The 1.2Å crystal structure of an *E. coli* tRNASer acceptor stem microhelix reveals two magnesium binding sites. *Biochem. Biophys. Res. Commun.* **2009**, *386*, 368−373.
(71) Popenda, L.; Adamiak, R. W.; Gdaniec, Z. Bulged Adenosine Influence on the RNA Duplex Conformation in Solution. *Biochemistry* **2008**, *47*, 5059−5067.
(72) Hays, F. A.; Teegarden, A.; Jones, Z. J. R.; Harms, M.; Raup, D.; Watson, J.; Cavaliere, E.; Ho, P. S. How sequence defines structure: A crystallographic map of DNA structure and conformation. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7157−7162.
(73) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2179−2183.
(74) Dauter, Z.; Adamiak, D. A. Anomalous signal of phosphorus used for phasing DNA oligomer: importance of data redundancy. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *57*, 990−995.
(75) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(76) Kopka, M. L.; Lavelle, L.; Han, G. W.; Ng, H.-L.; Dickerson, R. E. An Unusual Sugar Conformation in the Structure of an RNA/DNA Decamer of the Polypurine Tract May Affect Recognition by RNase H. *J. Mol. Biol.* **2003**, *334*, 653−665.
(77) Mohamed, A. A.; Wang, P. Y.; Bartel, D. P.; Vos, S. M. The structural basis for RNA slicing by human Argonaute2. *Cell Rep.* **2025**, *44*, No. 115166.
(78) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.
(79) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296−W303.
(80) Haider, S.; Parkinson, G. N.; Neidle, S. Crystal Structure of the Potassium Form of an Oxytricha nova G-quadruplex. *J. Mol. Biol.* **2002**, *320*, 189−200.
(81) Condon, D. E.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Yildirim, I.; Turner, D. H. Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 2729−2742.
(82) Abdelkafi, M.; Ghomi, M.; Turpin, P. Y.; Baumruk, V.; du Penhoat, C. H.; Lampire, O.; Bouchemal-Chibani, N.; Goyer, P.; Namane, A.; Gouyette, C.; Huynh-Dinh, T.; Bednárová, L. Common Structural Features of UUCG and UACG Tetraloops in Very Short Hairpins Determined by UV Absorption, Raman, IR and NMR Spectroscopies. *J. Biomol. Struct. Dyn.* **1997**, *14*, 579−593.
(83) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. J. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348−362.
(84) Xiong, Y.; Sundaralingam, M. Crystal structure of a DNA·RNA hybrid duplex with a polypurine RNA r(gaagaagag) and a complementary polypyrimidine DNA d(CTCTTCTTC). *Nucleic Acids Res.* **2000**, *28*, 2171−2176.
(85) Mráziková, K.; Mlýnský, V.; Kührová, P.; Pokorná, P.; Kruse, H.; Krepl, M.; Otyepka, M.; Banáš, P.; Šponer, J. UUCG RNA Tetraloop as a Formidable Force-Field Challenge for MD Simulations. *J. Chem. Theory Comput.* **2020**, *16*, 7601−7617.
(86) Love, O.; Galindo-Murillo, R.; Roe, D. R.; Dans, P. D.; Cheatham, T. E., III; Bergonzo, C. modXNA: A Modular Approach to Parametrization of Modified Nucleic Acids for Use with Amber Force Fields. *J. Chem. Theory Comput.* **2024**, *20*, 9354−9363.

https://doi.org/10.1021/acs.jctc.5c00602
J. Chem. Theory Comput. 2025, 21, 7535−7549